



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
-----------------	-------------	----------------------	---------------------	------------------

09/910,169

07/20/2001

Michel Decary

2937.1000-005

1867

21005

7590

06/24/2005

HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
530 VIRGINIA ROAD  
P.O. BOX 9133  
CONCORD, MA 01742-9133

EXAMINER

RIVERO, MINERVA

ART UNIT

PAPER NUMBER

2655

DATE MAILED: 06/24/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

# Office Action Summary

Application No.

09/910,169

Applicant(s)

DECARY ET AL.

Examiner

Minerva Rivero

Art Unit

2655

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

## Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

## Status

- 1) ☐ Responsive to communication(s) filed on \_\_\_\_.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

## Disposition of Claims

- 4) ☒ Claim(s) 1-48 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-48 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_ are subject to restriction and/or election requirement.

## Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 20 July 2001 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

## Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some \* c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
  - ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_.
  - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

## Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)  
Paper No(s)/Mail Date 1/16/03, 2/21/03, 3/21/03, 4/13/03, 5/6/03, 5/29/04, 4/12/04, 6/14/04, 8/30/04, 1/14/05
- 4) ☐ Interview Summary (PTO-413)  
Paper No(s)/Mail Date. \_\_\_\_
- 5) ☐ Notice of Informal Patent Application (PTO-152)
- 6) ☐ Other: \_\_\_\_

## DETAILED ACTION

### *Claim Rejections - 35 USC § 102*

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(a) the invention was known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the invention thereof by the applicant for a patent.

2. Claims 1, 3-5, 7, 9-11, 13-24, 26-31, 33-41, 44-46 and 48 are rejected under 35 U.S.C. 102(a) as being anticipated by Paik *et al.* (US Patent 6,076,088).

3. Regarding claim 38, Paik *et al.* disclose a method and computer apparatus for extracting information from a Web page comprising:

a source of Web pages of interest (*World Wide Web*, Col. 4, Lines 15-19);

an extractor coupled to receive Web pages from the source, the extractor being computer implemented and using natural language processing to extract desired information from the Web pages (*computer system for extracting information using natural language processing*, Col. 4, Lines 57-67); and

a storage subsystem coupled to the extractor for storing the extracted desired information in a data store (*file storage subsystem*, Col. 6, Line 67 – Col. 7, Line 2).

Art Unit: 2655

4. Regarding claims 1 and 39, Paik *et al.* disclose a method for extracting data from a Web page document comprising:

using natural language processing, finding possible formal names on a given Web page, the step of finding producing a first found set of formal names (*extracting information using natural language processing techniques*, Col. 4, Lines 57-66; *identifying proper names*, Col. 9, Lines 50-51; *the first concept in a CRC (concept-relation-concept) is a proper name*, Col. 3, Lines 49-58);

searching the given Web page for formal names not found by the natural language processing step of finding, said searching producing a second set of formal names (*using linguistic patterns to search and extract information*, Col. 11, Lines 25-38; *using pre-specified rule patterns to extract proper names*, Col. 16, Lines 44-59) [See Applicant's Specification, Page 8, Lines 1-3]; and

refining a combined set of formal names formed of the first found set and the second set, said refining producing a working set of people and organization names extracted from the given Web page (*removing redundant CRCs (concept-relation-concept)*, see *CRC Combiner*, Col. 20, Lines 22-30; *the first concept in a CRC (concept-relation-concept) is a proper name*, Col. 3, Lines 49-58).

5. Regarding claim 15, Paik *et al.* disclose a method for extracting information from a Web page document comprising:

performing a lexical analysis on a given Web page document to identify elements of interest, the elements of interest producing formal names (*extracting information*

Art Unit: 2655

*using natural language processing techniques*, Col. 4, Lines 57-66; *identifying proper names*, Col. 9, Lines 50-51; *the first concept in a CRC (concept-relation-concept) is a proper name*, Col. 3, Lines 49-58);

detecting a regular recurrence of a certain type of element, the detecting producing additional formal names (*using linguistic patterns to search and extract information*, Col. 11, Lines 25-38; *using pre-specified rule patterns to extract proper names*, Col. 16, Lines 44-59) [See Applicant's Specification, Page 8, Lines 1-3]);

resolving aliases of the produced formal names and additional formal names to form a working set of names of people and/or organizations named in the given Web page document (*removing redundant CRCs (concept-relation-concept)*, see *CRC Combiner*, Col. 20, Lines 22-30; *the first concept in a CRC (concept-relation-concept) is a proper name*, Col. 3, Lines 49-58).

6. Regarding claims 3 and 40, Paik *et al.* further disclose the step of and apparatus for refining includes determining aliases of respective people and organization names in the combined set, so as to reduce effective duplicate names (*using the standard form of a name for categorization, (thus avoiding forming various profiles for one individual)*, Col. 11, Lines 56-59).

7. Regarding claims 4 and 41, Paik *et al.* disclose the step of and apparatus for finding further finds professional titles and determines organization for which a person

Art Unit: 2655

named on the given Web page holds that title (Col. 20, Lines 31-38; *also see Company and Title under proper name categories in Table 1 and Name and Title in Table 9.2*).

8. Regarding claim 5, Paik *et al.* disclose the step of finding includes employing rules to extract at least title and formal names (*CRC Extractor is rule-based*, Col. 9, Lines 52-55).

9. Regarding claim 7, Paik *et al.* disclose the step of finding further includes determining biographical information relating to a person named on the given Web page (*creating an instant biography*, Col. 4, Line 63 – Col. 5, Line 6).

10. Regarding claim 9, Paik *et al.* disclose  
determining type of the given Web page (*identifying sentence and paragraph boundaries*, Col. 9, Lines 44-51; *identifying fields*, Col. 10, Lines 57-63); and  
from the determined type, defining contents of different portions of the Web page, such that the steps of finding and searching are performed as a function of the defined contents (*the identification process is fundamental to later natural language processing*, Col. 10, Lines 60-63).

11. Regarding claim 10, Paik *et al.* disclose the step of determining type of the given Web page includes determining structure or arrangements of contents of the page

*(identifying sentence and paragraph boundaries, Col. 9, Lines 44-51; identifying fields, Col. 10, Lines 57-63).*

12. Regarding claims 11 and 46, disclose the step of and apparatus for using the determined type, deducing additional information regarding a named person or organization on the given Web page, the additional information supplementing information found on another Web page of a same Web site as the given Web page *(retrieving all the information concerning a named entity, Col. 10, Lines 9-11).*

13. Regarding claim 13, Paik *et al.* disclose the step of searching employs pattern matching *(using linguistic patterns to search and extract information, Col. 11, Lines 25-38; using pre-specified rule patterns to extract proper names, Col. 16, Lines 44-59)* [See Applicant's Specification, Page 8, Lines 1-3].

14. Regarding claim 14, Paik *et al.* disclose a database having records formed by data extracted from Web pages *(acquiring new knowledge and adding it to the knowledge base, Col. 3, Lines 42-48).*

15. Regarding claim 16, Paik *et al.* disclose the step of transforming the given Web page document into a standardized form, the step of transforming including identifying page structure of the Web page document *(structurally parsing the documents and identifying sentence and paragraph boundaries, Col. 9, Lines 46-50).*

Art Unit: 2655

16. Regarding claim 17, Paik *et al.* disclose the step of assigning a type to each line in the given Web page document, the step of assigning a type indicating purpose of each line in the given Web page document (*preprocessing tags for identifying the various fields*, Col. 10, Lines 54-64). [See Applicant's Specification, Page 11, Lines 11-15.]

17. Regarding claim 18, Paik *et al.* disclose the step of performing a lexical analysis further identifies elements of interest on lines of certain assigned types (*further identifying fields and clauses in the text*, Col. 10, Lines 57-64).

18. Regarding claim 19, Paik *et al.* disclose the step of detecting using pattern matching, detecting a regular recurrence of a certain type of line, to produce additional formal names (*CRC extraction rules*, Col. 17, Lines 21-49).

19. Regarding claim 20, Paik *et al.* disclose the step of performing a lexical analysis includes syntactically and grammatically identifying elements of interest (*documents are parsed by a syntactic parser and tagged for parts of speech*, Col. 9, Lines 44-50).

20. Regarding claim 21, Paik *et al.* disclose the step of identifying elements of interest identifies noun phrases that correspond to a person or organization named in the given Web page document (*Proper Name Interpreter*, Col. 11, Lines 39-62).



Art Unit: 2655

21. Regarding claim 22, Paik *et al.* disclose the step of performing a lexical analysis includes using natural language processing (*extracting information using natural language processing techniques*, Col. 4, Lines 57-66; Col. 8, Lines 51-55).

22. Regarding claim 23, Paik *et al.* disclose the step of performing a lexical analysis includes utilizing rules describing composition of a name (*examining suffixes, prefixes and infixes*, Col. 11, Lines 53-55; *Proper Name Categories table*, Col. 26, Line 15 – Col. 27, Line 12). [See Applicant's Specification, Page 14, Lines 5-17.]

23. Regarding claim 24, Paik *et al.* disclose the step of resolving aliases includes employing rules for determining variant versions of a person's name or an organization's name (*proper name is passed to a database to determine if an alternative form exists*, Col. 11, Lines 56-59).

24. Regarding claim 26, Paik *et al.* disclose  
grouping subsets of lines together to form respective text units (*discourse-level decomposition of text*, Col. 10, Lines 51-63; *see also appositional phrases*, Col. 11, Line 67 – Col. 12, Line 3; *determining the boundaries between unique concepts*, Col. 12, Lines 29-34; *forming a single concept cluster*, Col. 12, Lines 65-66); and  
extracting from the formed text units desired information relating to the people or organizations named in the Web page document (*CRC (concept-relation-concept) triples*), Col. 13, Lines 57-67);

wherein the step of grouping identifies boundaries where information about a person or organization is to be found (*discourse-level decomposition of text*, Col. 10, Lines 51-63; *determining the boundaries between unique concepts*, Col. 12, Lines 29-34).

25. Regarding claim 27, Paik *et al.* suggest the step of grouping recognizes elements of information that span across more than one line (*discourse-level manipulation of text*, Col. 10, Lines 51-54).

26. Regarding claims 28 and 45, Paik *et al.* disclose the step of and apparatus for determining type of the given Web page (*identifying sentence and paragraph boundaries*, Col. 9, Lines 44-51; *identifying fields*, Col. 10, Lines 57-63); and from the determined type, defining contents of different portions of the Web page, such that the steps of finding and searching are performed as a function of the defined contents (*the identification process is fundamental to later natural language processing*, Col. 10, Lines 60-63).

27. Regarding claim 29, Paik *et al.* disclose the step of determining type of the given Web page includes determining structure or arrangements of contents of the page (*identifying sentence and paragraph boundaries*, Col. 9, Lines 44-51; *identifying fields*, Col. 10, Lines 57-63).

Art Unit: 2655

28. Regarding claim 30, Paik *et al.* disclose the step of extracting includes determining whether the given Web page document is a press release, and if so, identifying organization mentioned in the press release (*extracting all named entities and related information from news articles and news feeds and merging into a single profile*, Col. 4, Line 63 – Col. 5, Line 6).

29. Regarding claim 31, Paik *et al.* disclose the step of extracting includes using a parser to recognize the relationship between elements of information (*syntactic parser*, Col. 17, Line 61 – Col. 18, Line 12).

30. Regarding claim 33, Paik *et al.* suggest the step of extracting includes associating a person or organization with an element of information if said element appears in a non-sentence within a formed text unit for that person or organization (*extracting information using linguistic constructions in close proximity to a named entity and merging separate facts*, Col. 3, Line 59 – Col. 4, Line 5).

31. Regarding claim 34, Paik *et al.* disclose the step of extracting further divides a line that contains multiple names (*original sentence*, Col. 18, Lines 13-16; *CRC extractions (one for each person)*, Col. 18, Lines 58-61).

32. Regarding claims 35 and 44, Paik *et al.* disclose the step of and apparatus for extracting is rules-based (*CRC Extractor is rule-based*, Col. 9, Lines 52-55).

Art Unit: 2655

33. Regarding claims 36 and 48, Paik *et al.* disclose the step of and apparatus for post-processing to extract further names of organizations and relationships to people named in the given Web page document (*using four different CRC extraction modules and appropriately combining the output*, Col. 20, Lines 22-31).

34. Regarding claim 37, Paik *et al.* disclose the step of post-processing includes:  
extracting organization names from professional titles held by a named person (*noting affiliation between a named individual and a company*, Col. 18, Lines 13-16 and 58-61);

associating a named person with an organization whose Web site is hosting the given Web page document (*extracting information about all named entities and relation to any other named entity (persons, organizations, etc.) and merging it into a single profile with reference to the original sources*, Col. 4, Line 63 – Col. 5, Line 6); and

deducing organization names from biographical text of a named person (*parsing, tagging and CRC creation of a portion of an article*, Col. 14, Line 47 – Col. 15, Line 30).

### ***Claim Rejections - 35 USC § 103***

35. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the

Art Unit: 2655

invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

36. Claims 2 and 25 are rejected under 35 U.S.C. 103(a) as being unpatentable over Paik *et al.* (US Patent 6,076,088) in view of Asija (US Patent 4,270,182).

Regarding claims 2 and 25, Paik *et al.* do not disclose but Asija suggests the step of refining includes rejecting predefined formal names as not being people names of interest and rejecting names containing predefined forms of common known phrases (*listing common words, comparing them with text and discarding words that match*; Col. 3, Lines 15-30).

Therefore it would have been obvious to one ordinarily skilled in the art at the time of the invention to supplement the teachings of Paik *et al.* with rejecting predefined formal names as not being people names of interest and rejecting names containing predefined forms of common known phrases, as suggested by Asija, in order to reduce the text to that which contributes relevant information.

37. Claims 6, 8, 32 and 42-43 are rejected under 35 U.S.C. 103(a) as being unpatentable over Paik *et al.* (US Patent 6,076,088), as applied to claims 1 and 7 above, in view of Brady *et al.* (US Patent 6,463,430).

38. Regarding claims 6 and 42, Brady *et al.* disclose the step of and apparatus for finding includes determining educational background of a person named on the given Web page, the educational background including at least one of name of institution, degree earned from the institution and date of graduation from the institution (*academic training of an individual*, Col. 6, Lines 1-15; *extracting educational background information from Web documents*, Col. 19, Lines 24-38).

Therefore it would have been obvious to one ordinarily skilled in the art at the time of the invention to supplement the teachings of Paik *et al.* by having the step of finding includes determining educational background of a person named on the given Web page, the educational background including at least one of name of institution, degree earned from the institution and date of graduation from the institution, as taught by Brady *et al.*, in order to have a complete profile of an individual and accurately assess a candidate's aptitude for a particular job opportunity.

39. Regarding claims 8 and 43, Paik *et al.* do not explicitly disclose but Brady *et al.* do disclose the step of and apparatus for determining biographical information includes determining current and previous employment history of the named person (*employment experience*, Col. 6, Lines 1-15; *relevant experience*, Col. 19, Lines 24-38).

Therefore it would have been obvious to one ordinarily skilled in the art at the time of the invention to supplement the teachings of Paik *et al.* by having the step of determining biographical information includes determining current and previous employment history of the named person, as taught by Brady *et al.*, in order to have a

complete profile of an individual and accurately assess a candidate's aptitude for a particular job opportunity.

40. Regarding claim 32, Paik *et al.* disclose the step of extracting further includes utilizing predefined semantic frames for determining (i) sentences that express a relationship between a person and organization named in the given Web page document (*apposition identifiers, Apposition Evidence Database contains specific linguistic patterns*, Col. 11, Lines 25-38; *rule-based detection and extraction module, "A is a B at C of D"*, Col. 15, Line 60 – Col. 16, Line 10; specific relation extraction rules, Col. 16, Lines 24-44; *"entity has name" and "name has title"*, in *Table 2: Relations*, Col. 28, Lines 9-10) [See Applicant's Specification, Page 15, Lines 5-14.].

However, Paik *et al.* do not explicitly disclose but do Brady *et al.* suggest utilizing predefined semantic frames for determining sentences that express a person has a certain level of education (*academic training of an individual*, Col. 6, Lines 1-15; *extracting educational background information from Web documents*, Col. 19, Lines 24-38). Brady *et al.* disclose extracting educational background information of an individual.

Therefore it would have been obvious to one ordinarily skilled in the art at the time of the invention to supplement the teachings of Paik *et al.* with utilizing predefined semantic frames for determining sentences that express a person has a certain level of education, as suggested by Brady *et al.*, in order to assemble a more complete profile of a person named in the document.

Art Unit: 2655

41. Claims 12 and 47 are rejected under 35 U.S.C. 103(a) as being unpatentable over Paik *et al.* (US Patent 6,076,088), as applied to claims 1 and 38, in view of Smith *et al.* (US Patent 6,052,693).

Regarding claims 12 and 47, Paik *et al.* do not explicitly disclose but Smith *et al.* suggest the step of and apparatus for finding further includes determining at least one of addresses, telephone number, and email address relating to a person or organization named on the given Web page (Col. 3, Line 65 – Col. 4, Line 8).

Therefore it would have been obvious to one ordinarily skilled in the art at the time of the invention to supplement the teachings of Paik *et al.* with the step of finding further includes determining at least one of addresses, telephone number, and email address relating to a person or organization named on the given Web page, as taught by Smith *et al.*, in order to compile a more complete profile regarding the person being searched and have the ability to contact the individual or organization for employment or advertisement purposes.

### ***Conclusion***

42. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Appelt *et al.* (US Patent 6,601,026) disclose a querying system wherein



information from various sources regarding a particular topic is merged into a single profile.

McGreevy (US Patent 6,697,793) discloses a method for generating possible phrases from a specific context database to use in later querying and searching of the database.

Arnold *et al.* (US Patent 6,745,161) disclose a method for recognizing linguistic patterns and building an event structure for each found pattern, then merging the structures into a single informational profile.

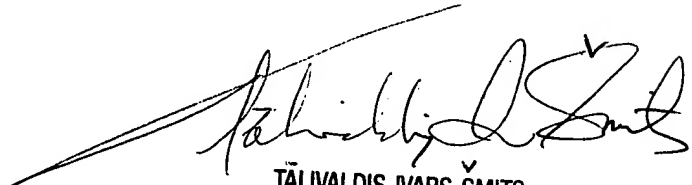
43. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Minerva Rivero whose telephone number is (571) 272-7626. The examiner can normally be reached on Monday-Friday 9:00 am - 6:00 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Talivaldis Ivars Smits can be reached on (571) 272-7628. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Art Unit: 2655

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

MR 6/21/2005



TĀLIVALDIS IVARS ŠMITS  
PRIMARY EXAMINER